B34

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12Q 1/70, C07K 5/00, 13/00
G01N 33/53

A1

(11) International Publication Number: WO 93/20242

(43) International Publication Date: 14 October 1993 (14.10.93)

(21) International Application Number: PCT/US93/03127

(22) International Filing Date: 30 March 1993 (30.03.93)

(30) Priority data:
07/860,445 30 March 1992 (30.03.92) US

(60) Parent Application or Grant
(63) Related by Continuation
US 07/860,445 (CIP)
Filed on 30 March 1992 (30.03.92)

(71) Applicant *(for all designated States except US)*: THE SCRIPPS RESEARCH INSTITUTE [US/US]; 10666 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : LERNER, Richard [US/US]; 7750 Roseland, La Jolla, CA 92037 (US). JANDA, Kim [US/US]; 3181 Erie Street, San Diego, CA 92107 (US). BRENNER, Sydney [GB/GB]; 17B Street, Edwards Passage, Cambridge CB2 3PJ (GB). NIELSEN, John [DK/US]; 5205 Fiore Terrace #B306, San Diego, CA 92122 (US).

(74) Agents: FITTING, Thomas et al.; The Scripps Research Institute, 10666 North Torrey Pines Road, TPC-8, La Jolla, CA 92037 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: ENCODED COMBINATORIAL CHEMICAL LIBRARIES

(57) Abstract

The present invention describes an encoded combinatorial chemical library comprised of a plurality of bifunctional molecules having both a chemical polymer and an identifier oligonucleotide sequence that defines the structure of the chemical polymer. Also described are the bifunctional molecules of the library, and methods of using the library to identify chemical structures within the library that bind to biologically active molecules in preselected binding interactions.

ENCODED COMBINATORIAL CHEMICAL LIBRARIES

Description

Technical Field

The present invention relates to encoded chemical libraries that contain repertoires of chemical structures defining a diversity of biological structures, and methods for using the libraries.

Background

There is an increasing need to find new molecules which can effectively modulate a wide range of biological processes, for applications in medicine and agriculture. A standard way for searching for novel bioactive chemicals is to screen collections of natural materials, such as fermentation broths or plant extracts, or libraries of synthesized molecules using assays which can range in complexity from simple binding reactions to elaborate physiological preparations. The screens often only provide leads which then require further improvement either by empirical methods or by chemical design. The process it time-consuming and costly but it is unlikely to be totally replaced by rational methods even when they are based on detailed knowledge of the chemical structure of the target molecules. Thus, what we might call "irrational drug design" - the process of selecting the right molecules from large ensembles or repertoires - requires continual improvement both in the generation of repertoires and in the methods of selection.

Recently there have been several developments in using peptides or nucleotides to provide libraries of compounds for lead discovery. The methods were originally developed to speed up the determination of

reiterated in this fashion until the active hexapeptide is defined. This is analogous to the method used in searching a dictionary; the peptide is decoded by construction using a series of sieves or buckets and this makes the search logarithmic.

A very powerful biological method has recently been described in which the library of peptides is presented on the surface of a bacteriophage such that each phage has an individual peptide and contains the DNA sequence specifying it. The library is made by synthesizing a repertoire of random oligonucleotides to generate all combinations, followed by their insertion into a phage vector. Each of the sequences is cloned in one phage and the relevant peptide can be selected by finding those that bind to the particular target. The phages recovered in this way can be amplified and the selection repeated. The sequence of the peptide is decoded by sequencing the DNA. See for example Cwirla et al., Proc.Natl.Acad.Sci.USA, 87:6378-6382 (1990); Scott et al., Science, 249:386-390 (1990); and Devlin et al., Science, 249:404-406 (1990).

Another "genetic" method has been described where the libraries are the synthetic oligonucleotides themselves wherein active oligonucleotide molecules are selected by binding to an acceptor and are then amplified by the polymerase chain reaction (PCR). PCR allows serial enrichment and the structure of the active molecules is then decoded by DNA sequencing on clones generated from the PCR products. The repertoire is limited to nucleotides and the natural pyrimidine and purine bases or those modifications that preserve specific Watson-Crick pairing and can be copied by polymerases.

The main advantages of the genetic methods reside

molecule of interest or by identifying a species in the library having a desired activity, including binding, activation, chemical catalysis and the like. Thereafter, the identity of the active molecule is determined by reading the genetic tag, i.e., the identifier oligonucleotide sequence. In one embodiment, amplified copies of their retrogenetic tags can be obtained by the polymerase chain reaction.

The strands of the amplified copies with the appropriate polarity can then be used to enrich for a subset of the library by hybridization with the matching tags and the process can then be repeated on this subset. Thus serial enrichment is achieved by a process of purification exploiting linkage to a nucleotide sequence which can be amplified. Finally, the structure of the chemical entities are decoded by cloning and sequencing the products of the PCR reaction.

The present invention therefore provides a novel method for identifying a chemical structure having a preselected binding or catalysis activity through the use of a library of bifunctional molecules that provides a rich source of chemical diversity. The library is used to identify chemical structures (structural motifs) that interact with preselected biological molecules.

Thus, in one embodiment, the invention contemplates a bifunctional molecule according to the formula A-B-C, where A is a chemical moiety, B is a linker molecule operatively linked to A and C, and C is an identifier oligonucleotide comprising a sequence of nucleotides that identifies the structure of chemical moiety A.

In another embodiment, the invention contemplates a library comprising a plurality of species of

aliquots of the composition to produce aliquots that contain a product containing a bifunctional molecule;

d) combining the aliquots produced in step (c) to form an admixture of bifunctional molecules, thereby forming said library.

In a related embodiment, the invention describes a bifunctional solid support for synthesizing an oligopeptide/ oligonucleotide conjugate comprising: (1) a solid support, the solid support being of a type which is dispersible in aqueous solution, (2) a first linkage unit coupled to the solid support, (3) a second linkage unit coupled to the first linkage unit, and (4) a bifunctional unit coupled to the second linkage unit, wherein the bifunctional unit having a first leaving group employable for oligopeptide synthesis and a second leaving group employable for oligonucleotide synthesis, wherein the first leaving group is N-FMOC or its functional equivalent, the second leaving group is O-DMT or its functional equivalent, wherein the second linkage unit is coupled to the first linkage unit by means of a bond cleavable by exposure to concentrated aqueous ammonia, and wherein the solid support, the first linkage unit, the second linkage unit, the cleavable bond, and the bifunctional unit, exclusive of the first and second leaving groups, each being substantially chemically unreactive to conditions employed by oligopeptide synthetic protocols using FMOC leaving groups and conditions employed by oligonucleotide synthetic protocols using O-DMT leaving groups.

Further contemplated is an element of a library of oligopeptide/oligonucleotide conjugates comprising: (1) a solid support being of a type which is dispersible in aqueous solution, (2) a first linkage unit coupled to the solid support, (3) a second

elements of a polymer comprised of a series of chemical building blocks to form a chemical moiety in the library, and a code for identifying the structure of the chemical moiety.

Thus, a bifunctional molecule can be represented by the formula A-B-C, where A is a chemical moiety, B is a linker molecule operatively linked to A and C, and C is an identifier oligonucleotide comprising a sequence of nucleotides that identifies the structure of chemical moiety A.

a. Chemical Polymers

A chemical moiety in a bifunctional molecule of this invention is represented by A in the above formula A-B-C and is a polymer comprising a linear series of chemical units represented by the formula $(X_n)_a$, wherein X is a single chemical unit in polymer A and n is a position identifier for X in polymer A. n has the value of 1+i where i is an integer from 0 to 10, such that when n is 1, X is located most proximal to the linker (B).

Although the length of the polymer can vary, defined by a, practical library size limitations arise if there is a large alphabet size as discussed further herein. Typically, a is an integer from 4 to 50.

A chemical moiety (polymer A) can be any of a variety of polymeric structures, depending on the choice of classes of chemical diversity to be represented in a library of this invention. Polymer A can be any monomeric chemical unit that can be coupled and extended in polymeric form. For example, polymer A can be a polypeptide, oligosaccharide, glycolipid, lipid, proteoglycan, glycopeptide, sulfonamide, nucleoprotein, conjugated peptide (i.e., having prosthetic groups), polymer containing enzyme

F	Phe	phenylalanine
M	Met	methionine
A	Ala	alanine
S	Ser	serine
I	Ile	isoleucine
L	Leu	leucine
T	Thr	threonine
V	Val	valine
P	Pro	proline
K	Lys	lysine
H	His	histidine
Q	Gln	glutamine
E	Glu	glutamic acid
W	Trp	tryptophan
R	Arg	arginine
D	Asp	aspartic acid
N	Asn	asparagine
C	Cys	cysteine

The phrase "amino acid residue" is broadly defined to include the amino acids listed in the Table of Correspondence and modified and unusual amino acids, such as those listed in 37 C.F.R. §1.822(b)(4), and incorporated herein by reference.

The polymer defined by chemical moiety A can therefore contain any polymer backbone modifications that provide increased chemical diversity. In building of a polypeptide system as exemplary, a variety of modifications are contemplated, including the following backbone structures: -NHN(R)CO-, -NHB(R)CO-, -NHC(RR')CO-, -NHC(=CHR)CO-, $-NHC_6H_4CO-$, $-NHCH_2CHRCO-$, $-NHCHRCH_2CO-$, and lactam structures.

In addition, amide bond modifications are contemplated including $-COCH_2-$, -COS-, -CONR, -COO-, -CSNH-, $-CH_2NH-$, $-CH_2CH_2-$, $-CH_2S-$, $-CH_2SO-$, $-CH_2SO_2-$,

Insofar as adenosine (A), guanosine (G), thymidine (T) and cytidine (C) represent the typical choices of deoxynucleotides for inclusion in a unit identifier oligonucleotide, A, G, T and C form a representative "alphabet" used to "spell" out a unit identifier oligonucleotide's sequence. Other nucleotides or nucleotide analogs can be utilized in addition to or in place of the above four nucleotides, so long as they have the ability to form Watson-Crick pairs and be replicated by DNA polymerases in a PCR reaction. However, the nucleotides A, G, T and C are preferred.

For the design of the code in the identifier oligonucleotide, it is essential to chose a coding representation such that no significant part of the oligonucleotide sequence can occur in another unrelated combination by chance or otherwise during the manipulations of a bifunctional molecule in the library.

For example, consider a library where Z is a trinucleotide whose sequence defines a unique chemical unit X. Because the methods of this invention provide for all combinations and permutations of an alphabet of chemical units, it is possible for two different unit identifier oligonucleotide sequences to have closely related sequences that differ by only a frame shift and therefore are not easily distinguishable by hybridization or sequencing unless the frame is clear.

Other sources of misreading of a unit identifier oligonucleotide can arise. For example, mismatch in DNA hybridization, transcription errors during a primer extension reaction to amplify or sequence the identifier oligonucleotide, and the like errors can occur during a manipulation of a bifunctional

discussed in more detail herein.

In a preferred embodiment, an identifier oligonucleotide C has a nucleotide sequence according to the formula P1-$(Z_n)_a$-P2, where P1 and P2 are nucleotide sequences that provide polymerase chain reaction (PCR) primer binding sites adapted to amplify the polymer identifier oligonucleotide. The requirements for PCR primer binding sites are generally well known in the art, but are designed to allow a PCR amplification product (a PCR-amplified duplex DNA fragment) to be formed that contains the polymer identifier oligonucleotide sequences.

The presence of the two PCR primer binding sites, P1 and P2, flanking the identifier oligonucleotide sequence $(Z_n)_a$ provides a means to produce a PCR-amplified duplex DNA fragment derived from the bifunctional molecule using PCR. This design is useful to allow the amplification of the tag sequence present on a particular bifunctional molecule for cloning and sequencing purposes in the process of reading the identifier code to determine the structure of the chemical moiety in the bifunctional molecule.

More preferred is a bifunctional molecule where one or both of the nucleotide sequences P1 and P2 are designed to contain a means for removing the PCR primer binding sites from the identifier oligonucleotide sequences. Removal of the flanking P1 and P2 sequences is desirable so that their sequences do not contribute to a subsequent hybridization reaction. Preferred means for removing the PCR primer binding sites from a PCR amplification product is in the form of a restriction endonuclease site within the PCR-amplified duplex DNA fragment.

Restriction endonucleases are well known in the art and are enzymes that recognize specific lengths of

termini. Step 2 illustrates the specific biotinylation of the anticoding strand at the Sty I site, whereby the incorporation of biotinylated UTP is indicated by a B.

The presence of non-overlapping cohesive termini after Step 1 in Figure 1 allows the specific and directional cloning of the restriction-digested PCR-amplified fragment into an appropriate vector, such as a sequencing vector. In addition, the Sty I was designed into P1 because the resulting overhang is a substrate for a filling-in reaction with dCTP and biotinyl-dUTP (BTP) using DNA polymerase Klenow fragment. The other restriction site, Apa I, was selected to not provide substrate for the above biotinylation, so that only the anticoding strand can be biotinylated.

Once biotinylated, the duplex fragment can be bound to immobilized avidin and the duplex can be denatured to release the coding sequence containing the identifier nucleotide sequence, thereby providing purified anticoding strand that is useful as a hybridization reagent for selection of related coding strands as described further herein.

c. Linker Molecules

A linker molecule in a bifunctional molecule of this invention is represented by B in the above formula A-B-C and can be any molecule that performs the function of operatively linking the chemical moiety to the identifier oligonucleotide.

Preferably, a linker molecule has a means for attaching to a solid support, thereby facilitating synthesis of the bifunctional molecule in the solid phase. In addition, attachment to a solid support

of a 2-hydroxymethyl benzoic acid moiety provides a site cleavable by the combination of alkaline phosphatase treatment followed by mild alkaline treatment. Thus, the incorporation of selectively cleavable linkers other than those recited above are also considered a part of the invention.

Solid supports for chemical synthesis are generally well known. Particularly preferred are the synthetic resins and controlled pore glass (CPG) supports used in oligonucleotide and in polypeptide synthesis that are available from a variety of commercial sources including Glen Research (Herndon, VA), Bachem Biosciences (Philadelphia, PA), Sigma Chemical Co. (St. Louis, MO), CPG Inc., (Fairfield, NJ) and Applied Biosystems (Foster City, CA). Most preferred are teflon and CPG supports such as are described in Example 2.

In a related embodiment, the invention describes a preferrred bifunctional solid support particularly suited for producing a bifunctional molecule of this invention having a polypeptide as the chemical moiety (i.e., an oligopeptide/oligonucleotide conjugate).

A preferred bifunctional solid support for synthesizing oligopeptide/oligonucleotide conjugates comprises a solid support, a first linkage unit, a second linkage unit, and a bifunctional unit. A "bifunctional unit" in this context is not to be confused with a bifunctional molecule of this invention, and refers instead to that chemical moiety present on a bifunctional solid support which provides the two (bi-) reactive functionalities, one for coupling the peptide and another for coupling the oligonucleotide. An exemplary bifunctional unit is shown in Figure 3 as the serine-branch monomer following the aminohexanol linker.

and the bifunctional unit, exclusive of said first and second leaving groups, are each substantially chemically unreactive under conditions employed during conventional oligopeptide synthesis protocols using FMOC leaving groups [see: Bodanszky et al., in *The practice of Peptide Synthesis*, Springer-Verlag, (1984); and Bodanszky et al., in *Principles of Peptide Synthesis*, Springer-Verlag, (1984)] and under conditions employed during conventional oligonucleotide synthesis protocols using O-DMT leaving groups and phosphoramidite donors.

In an alternative embodiment, the bifunctional solid support also includes a third linkage unit. The third linkage unit is interposed between and coupled to the bifunctional unit and the first leaving group. In a preferred mode of this embodiment, the third linkage unit is photosensitive, i.e. it is cleavable by exposure to ultra-violet light. The third linkage unit may include a 3-nitro-4-bromomethyl benzoate group coupled by means of an amide bond to the amino end of the bifunctional serine residue and coupled by means of an ester bond to an FMOC blocked amino acid.

2. Libraries

A library of this invention is a repertoire of chemical diversity comprising a plurality of species of bifunctional molecules according to the present invention. The plurality of species in a library defines a family of chemical diversity whose species each have a different chemical moiety. Thus the library can define a family of peptides, lipids, oligosaccarides or any of the other classes of chemical polymers recited previously.

The number of different species in a library represents the complexity of a library and is defined

molecule to be assayed and a bifunctional molecule.

If, for example, the binding threshold is 10^{-6}M (micromolar), then there must be at least one nanomole of each species in a library of 1 milliliter (ml) volume. At this threshold, a library having a complexity of 10^4 could contain 10 micromoles of each species. Because of the reciprocal relationship between library complexity and binding threshold, more complex libraries are possible where the binding threshold is lower.

The relative amounts of the individual bifunctional molecule species within the library can vary from about 0.2 equivalents to about 10 equivalents, where an equivalent represents the average amount of a species within the library. Preferably each species is present in the library in approximately equimolar amounts.

In a preferred embodiment, a library contains the complete repertoire of chemical diversity possible based on the mathematical combinations for a given library where there is a fixed alphabet and a preselected number of chemical units in all species of the library. Thus a complete repertoire is one that provides a source of all the possible chemical diversity that can be found in a library of this invention having a fixed alphabet and chemical length.

It is particularly preferred that a library be comprised of bifunctional molecules where each species of bifunctional molecule contains the same nucleotide sequence for either the P1 or P2 PCR primer binding sites. A library with this design is particularly preferred because, when practicing the methods of this invention, a single PCR primer pair can be used to amplify any species of identifier oligonucleotide (coding sequence) present in the library.

attached to said bifunctional unit, and an oligonucleotide attached to said bifunctional unit. An alternative to this embodiment includes a cleavable bond for coupling the bifunctional unit to the oligopeptide. The cleavable bond may be of a type which is photosensitive, i.e. cleavable by exposure to ultraviolet light, as indicated above.

B. Methods for Producing a Library

The present method for producing a plurality of bifunctional molecules to form a library of this invention solves a variety of problems regarding efficient synthesis of large numbers of different species.

In the present synthesis methods, the sequential steps of first adding a chemical unit X followed by the addition of an oligonucleotide sequence to the linker molecule requires an alternating parallel synthesis procedure to add chemical unit X and then add a unit identifier nucleotide sequence Z that defines (codes for) that corresponding chemical unit. The library is built up by the repetition of this alternating parallel process after pooling and division of the reaction products as described herein.

The only constraint for making an encoded library is that there must be compatible chemistries between the two alternating syntheses procedures for adding a chemical unit as compared to that for adding a nucleotide or oligonucleotide sequence.

The problem of synthesis compatibility is solved by the correct choice of compatible protecting groups as the alternating polymers are synthesized, and by the correct choice of methods for deprotection of one growing polymer selectively while the other growing polymer remains blocked, such as by the use of

identifier oligonucleotide Z first to termini C'. A first cycle involves the steps of deprotecting the termini of the linker to which a building block is to be added and then adding the building block to the termini. Typically, the added building block contains a blocking group at its free termini, i.e., the termini that will participate in an addition of the next building block of its type. The linker molecule is then subjected to a second cycle of synthesis to add a building block at the other (second) termini. A second cycle involves the steps of deprotecting the second termini of the linker to which a building block is to be added and then adding the building block to the termini. Again, the added building block is typically blocked at its free termini.

The addition of identifier oligonucleotide Z to termini C' can be conducted either nucleotide by nucleotide to form the complete unit identifier nucleotide sequence Z, or Z can be presynthesized, and the oligonucleotide Z added as a block to termini C'. Insofar as the synthesis of oligonucleotides is well known in the arts, the presynthesis of oligonucleotides, and their addition to the growing nucleotide polymer in blocks is preferred because it reduces the number of manipulations in synthesizing a bifunctional molecule.

A chemical unit X or a unit identifier oligonucleotide Z is referred to as a precursor (X' or Z') to indicate that it contains a leaving group compatible with the reaction chemistry that facilitates the precursor's operative linkage to the growing polymer at the appropriate termini.

The product resulting from step (2) is a bifunctional molecule having the structure A'-X_1-B-Z_1-C', and is ready for a repetition of the above first

Thus the steps of (i) dividing a linker or pool into aliquots, (ii) parallel addition of X and Z to the linker substrate in separate aliquots, and (iii) pooling of the aliquots, can be cycled (repeated) to sequentially add the chemical units and their corresponding unit identifier oligonucleotides to form the library comprising a plurality of bifunctional molecules each having a different chemical polymer operatively linked through the linker to a corresponding identifier oligonucleotide.

In a preferred embodiment, a method for forming a library of this invention includes the steps for addition of the PCR primer binding sites P1 and P2 to each of the bifunctional molecules in the library.

The method is substantially the same as above, but includes the addition of a series of nucleotides or a presynthesized P1 oligonucleotide to the linker molecule provided in step (1) prior to the cycles of step (2) that add X and Z. Because all members of the library are to contain the same P1 sequence, P1 is added to the C' termini of linker molecule A'-B-C' prior to dividing the linker into aliquots and subjecting the aliquots to the cycles of step (2) adding X_1 and Z_1. The resulting product has the formula A'-B-P1-C'.

Thereafter, the product is aliquoted and cycled as before, resulting in the preparation of the product $A'-(X_n)_a-B-P1-(Z_n)_a-C'$, where a indicates the presence of a polymer of length "a".

Next, the pooled admixture containing product $A'-(X_n)_a-B-P1-(Z_n)_a-C'$ is subjected to the addition of a series of nucleotides or a presynthesized oligonucleotide P2 at termini C' to form the product $A'-(X_n)_a-B-P1-(Z_n)_a-P2-C'$. Thus all members of the library contain a common sequence P1 and a common

to any particular blocking group. Also indicated in Table 1 are preferred blocking groups.

TABLE 1

Amino Acid	Blocking Group
Arginine	N-MTr[1], N-PMC[7]
Histidine	N^{π}-Bum[2], SEM[10], FMOC, DNP
Cysteine	S-Trt[3], Acm[11], S-t-butyl
Tryptophan	N^{i}-CHO, none
Tyrosine	O-TBS[4]
Aspartic acid	O-TSE[5], DMB[12]
Glutamic acid	O-TSE[5], DMB
Serine	O-TBS[4]
Threonine	O-TBS[4]
Lysine	N-Bz[6], TFA[8], TEOC[9]
Asparagine	none
Glutamine	none
Glycine	none
Phenylalanine	none
Methionine	none
Alanine	none
Isoleucine	none
Leucine	none
Valine	none
Proline	none

1 MTr is N^{9}-4-methoxy-2,3,6-trimethylbenzene sulfonyl.

2 Bum is tert-butoxymethyl.

3 Trt is triphenylmethyl.

5 TSE is trimethylsylilethylester.

4 TBS is tert-butyl-dimethylsilylester.

6 Bz is benzyl.

7 PMC is N_{G}-2,2,5,7,8-pentamethylchromon-6-

yl)-1,1,3,3-tetramethyluronium hexafluorophosphate. TBTU is 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate. pyBOP is bensotriazol-1-yloxytris[dimethylamino phosphonium hexafluorophosphate. The addition reaction requires the blocked amino acid, dimethylformamide (DMF) and hydroxy-benzotriazole (HOBt) as is well known for peptide synthesis. The resulting product contains an added amino acid residue with a FMOC-blocked amino terminus, ready for deblocking addition of a subsequent blocked amino acid as before .

For synthesis of a polypeptide on the linker substrate in the direction of amino to carboxy terminus, a free carboxy terminus on the linker is required that can be conveniently blocked and deblocked as needed. A preferred carboxy terminus blocking/activating group is the Opfp ester described before. A carboxy terminus on the linker is produced by reacting a linker with a free amino terminus with succinamide in HOBT and a proton catalyst. Thereafter, the terminus can be modified by reaction with pentafluorophenol in dicyclohexylcarbodiimide (DCC) and ethanol acetate to form an Opfp ester at the free carboxy terminus. The Opfp ester is blocked linker terminus is available for addition reaction with a FMOC-, Opfp- blocked amino acid as before, but with the amino acid adding to the linker in the reverse direction. The resulting product contains an added amino acid residue with an Opfp-blocked terminus, ready to repeat the addition with a subsequent blocked amino acid.

b. Oligonucleotide Synthesis

Oligonucleotides can be synthesized by a variety of chemistries as is well

ready for deblocking and addition of a subsequent blocked nucleotide as before.

For synthesis of an oligonucleotide on the linker in the direction of 5' to 3', a free hydroxy terminus on the linker is required as before. However, the blocked nucleotide to be added has the blocking chemistries reversed on its 5' and 3' termini to facilitate addition in the opposite orientation.

A nucleotide with a free 3' hydroxyl and 5' DMT ether is first blocked at the 3' hydroxy terminus by reaction with TBS-Cl in imidazole to form a TBS ester at the 3' terminus. Then the DMT- blocked 5' terminus is deblocked with DCA in DCM as before to form a free 5' hydroxy terminus. The reagent (N,N-diisopropylamino)(cyanoethyl) phosphonamidic chloride is reacted in tetrahydrofuran (THF) with the 5' deblocked nucleotide to form the aminodiisopropyl-, CNE- blocked phosphoramidite group on the 5' terminus. Thereafter the 3' TBS ester is removed with tetrabutylammonium fluoride (TBAF) in DCM to form a nucleotide with the phosphoramidite-blocked 5' terminus and a free 3' hydroxy terminus. Reaction in base with DMT-Cl adds a DMT ether blocking group to the 3' hydroxy terminus.

The addition of the 3' DMT-, 5' CNE- blocked phosphoramidited nucleotide to a linker substrate having a free hydroxy terminus then proceeds using the previous tetrazole catalyzed reaction, as is well known for oligonucleotide synthesis. The resulting product contains an added nucleotide residue with a DMT-blocked 3' terminus, ready for deblocking with DCA in DCM and the addition of a subsequent blocked nucleotide as before.

The above demonstrates that the present bifunctional molecules can be synthesized having

In the present preferred embodiment for a polypeptide library, the following sequence of deprotecting is preferred when using the teflon support and 5'BCM3 linker described in Examples 2-3:

1) tetrabutyl ammonium fluoride (TBAF) treatment to remove TBS and TMS ethyl ethers;

2) brief 5 minute treatment with trifluoroacetic acid (TFA) treatment sufficient to remove MTr, Bum, PMC and Trt groups, followed by neutralization for 5 minutes with triethylamine;

3) aqueous ammonia treatments to remove Bz and CNE groups; and

4) cleavage of the bifunctional molecule from the solid support using a periodate oxidation.

Alternatively, the following sequence of deprotection is used on the CPG linker described in Example 3 either with or without the photoactively cleavable linker moiety:

1) TBAF treatment to remove TBS and TMS ethyl ethers; and

2) aqueous ammonia treatment to remove Bz and CNE groups.

As indicated, after the library has been synthesized, and after the protecting groups have been removed, the bifunctional molecules may be cleaved off of the solid support, and the released bifunctional molecules separated from the solid phase to form a solution comprising a plurality of bifunctional molecules. Alternatively, the library may be maintained in the form of a plurality of bifunctional molecules in the solid phase.

Although natural amino acids are used in the Examples, the present invention is not to be so limited. The alphabet of possible amino acid residues can be extended to include any molecule that satisfies

facilitates identification of the chemical structure.

By the present screening methods, one can identify optimized chemical structures that participate in binding interactions or chemical catalysis events with a biologically active molecule by drawing upon a repertoire of structures randomly formed by the combinatorial association of diverse chemical units without the necessity of either synthesizing them one at a time or knowing their interactions in advance.

The invention therefore also contemplates a method for identifying a chemical structure that participates in a preselected binding or catalysis interactions between the chemical structure and a biologically active molecule. The chemical structure to be identified is represented by one of the members of a library of this invention, and the method comprises the following steps:

(1) A library according to the present invention is admixed with a preselected biologically active molecule under binding conditions (i.e., a binding reaction admixture) for a time period sufficient for the biologically active molecule to interact with at least one bifunctional molecule of this invention present in the library and form a binding reaction complex.

(2) The binding reaction complex is then isolated from the library admixture to form an isolated complex.

(3) The nucleotide sequence of the polymer identifier oligonucleotide present in the isolated binding reaction complex is determined. The nucleotide sequence provides a code that defines the chemical structure that participated in the binding reaction, and thus determining that sequence

phase. Still further, both the library members and the biologically active molecule can be in the liquid phase.

Binding conditions are those conditions compatible with the known natural binding function of the biologically active molecule. Those compatible conditions are buffer, pH and temperature conditions that maintain the biological activity of the biologically active molecule, thereby maintaining the ability of the molecule to participate in its preselected binding interaction. Typically, those conditions include an aqueous, physiologic solution of pH and ionic strength normally associated with the biologically active molecule of interest.

For example, where the binding interaction is to identify a member in the library able to bind an antibody molecule, the preferred binding conditions would be conditions suitable for the antibody to immunoreact with its immunogen, or a known immunoreacting antigen. For a receptor molecule, the binding conditions would be those compatible with measuring receptor-ligand interactions.

A time period sufficient for the admixture to form a binding reaction complex is typically that length of time required for the biologically active molecule to interact with its normal binding partner under conditions compatible with interaction. Although the time periods can vary depending on the molecule, admixing times are typically for at least a few minutes, and usually not longer than several hours, although nothing is to preclude using longer admixing times for a binding reaction complex to form.

A binding reaction complex is a stable product of the interaction between a biologically active molecule and a bifunctional molecule of this invention. The

immunoreacts with the biologically active molecule present in the binding reaction admixture to form an antibody-biologically active molecule immunoreaction complex. Thereafter, by separation of the solid phase from the binding reaction admixture, the immunoreaction complex, and therefor any binding reaction complex, is separated from the admixture to form isolated bifunctional molecule.

Alternatively, a binding means can be operatively linked to the biologically active molecule to facilitate its retrieval from the binding reaction admixture. Exemplary binding means are one of the following high affinity pairs: biotin-avidin, protein A-Fc receptor, ferritin-magnetic beads, and the like. Thus, the biologically active molecule is operatively linked (conjugated) to biotin, protein A, ferritin and the like binding means, and the binding reaction complex is isolated by the use of the corresponding binding partner in the solid phase, e.g., solid-phase avidin, solid-phase Fc receptor, solid phase magnetic beads and the like.

The use of solid supports on which to operatively link proteinaceous molecules is generally well known in the art. Useful solid support matrices are well known in the art and include cross-linked dextran such as that available under the tradename SEPHADEX from Pharmacia Fine Chemicals (Piscataway, NJ); agarose, borosilicate, polystyrene or latex beads about 1 micron to about 5 millimeters in diameter, polyvinyl chloride, polystyrene, cross-linked polyacrylamide, nitrocellulose or nylon-based webs such as sheets, strips, paddles, plates microtiter plate wells and the like insoluble matrices.

the amplified fragments into sequencing vectors. The cloning and sequencing of the amplified fragments then is a routine procedure that can be carried out by any of a number of molecular biological methods known in the art.

Alternatively, PCR amplified products derived from a population of isolated bifunctional molecules can be used as a hybridization probe to selectively enrich the quality of the isolated bifunctional molecules. For example, using the hybridization probes, which are modified by biotinylation as shown in Figure 1, one can isolate members of the library by hybridization, to form an enriched library containing only bifunctional molecules that have sequences that hybridize to the above hybridization probes. In a second screening reaction under different binding conditions, for example, higher stringency binding conditions, one can isolate the species of bifunctional molecule that binds most tightly with the biologically active molecule.

Thus the library can be manipulated to form enriched libraries from which to screen for chemical diversity.

4. Polymerase Chain Reaction

For determining the nucleotide sequence of the identifier oligonucleotide in the isolated complex as part of the methods of this invention, the use of the polymerase chain reaction (PCR) is a preferred embodiment.

For use in this invention, the identifier oligonucleotide are comprised of polynucleotide coding strands, such as mRNA and/or the sense strand of genomic DNA. If the genetic material to be assayed is in the form of double stranded DNA, it is usually

phasic with time periods of relative temperature stability at each of temperatures favoring polynucleotide synthesis, denaturation and hybridization.

A plurality of first primer and/or a plurality of second primers can be used in each amplification, e.g., one species of first primer can be paired with a number of different second primers to form several different primer pairs. Alternatively, an individual pair of first and second primers can be used. In any case, the amplification products of amplifications using the same or different combinations of first and second primers can be combined for assaying for mutations.

The PCR reaction is performed using any suitable method. Generally it occurs in a buffered aqueous solution, i.e., a PCR buffer, preferably at a pH of 7-9, most preferably about 8. Preferably, a molar excess of the primer is admixed to the buffer containing the template strand. A large molar excess is preferred to improve the efficiency of the process.

The PCR buffer also contains the deoxyribonucleotide triphosphates (polynucleotide synthesis substrates) dATP, dCTP, dGTP, and dTTP and a polymerase, typically thermostable, all in adequate amounts for primer extension (polynucleotide synthesis) reaction. The resulting solution (PCR admixture) is heated to about 90°C - 100°C for about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period the solution is allowed to cool to 54°C, which is preferable for primer hybridization. The synthesis reaction may occur at from room temperature up to a temperature above which the polymerase (inducing agent) no longer functions efficiently. Thus, for example, if DNA polymerase is

polymerase amplifies the starting polynucleotide as has been described by Chamberlin et al., The Enzymes, ed. P. Boyer, pp. 87-108, Academic Press, New York (1982). Amplification systems based on transcription have been described by Gingeras et al., in PCR Protocols, A Guide to Methods and Applications, pp. 245-252, Innis et al., eds, Academic Press, Inc., San Diego, CA (1990).

If the inducing agent is a DNA-dependent RNA polymerase and, therefore incorporates ribonucleotide triphosphates, sufficient amounts of ATP, CTP, GTP and UTP are admixed to the primer extension reaction admixture and the resulting solution is treated as described above.

The newly synthesized strand and its complementary nucleic acid strand form a double-stranded molecule which can be used in the succeeding steps of the process.

PCR amplification methods are described in detail in U.S. Patent Nos. 4,683,192, 4,683,202, 4,800,159, and 4,965,188, and at least in several texts including PCR Technology: Principles and Applications for DNA Amplification, H. Erlich, ed., Stockton Press, New York (1989); and PCR Protocols: A Guide to Methods and Applications, Innis et al., eds., Academic Press, San Diego, California (1990).

The term "polynucleotide" as used herein in reference to primers, probes and nucleic acid fragments or segments to be synthesized by primer extension is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three. Its exact size will depend on many factors, which in turn depends on the ultimate conditions of use.

The term "primer" as used herein refers to a

non-complementary nucleotide fragment can be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Such non-complementary fragments typically code for an endonuclease restriction site. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided the primer sequence has sufficient complementarity with the sequence of the strand to be synthesized or amplified to non-randomly hybridize therewith and thereby form an extension product under polynucleotide synthesizing conditions.

Primers of the present invention may also contain a DNA-dependent RNA polymerase promoter sequence or its complement. See for example, Krieg et al., Nucl. Acids Res., 12:7057-70 (1984); Studier et al., J. Mol. Biol., 189:113-130 (1986); and Molecular Cloning: A Laboratory Manual, Second Edition, Maniatis et al., eds., Cold Spring Harbor, NY (1989).

When a primer containing a DNA-dependent RNA polymerase promoter is used, the primer is hybridized to the polynucleotide strand to be amplified and the second polynucleotide strand of the DNA-dependent RNA polymerase promoter is completed using an inducing agent such as E. coli DNA polymerase I, or the Klenow fragment of E. coli DNA polymerase. The starting polynucleotide is amplified by alternating between the production of an RNA polynucleotide and DNA polynucleotide.

Primers may also contain a template sequence or replication initiation site for a RNA-directed RNA polymerase. Typical RNA-directed RNA polymerase include the QB replicase described by Lizardi et al., Biotechnology, 6:1197-1202 (1988). RNA-directed polymerases produce large numbers of RNA strands from

synthesis, i.e., initiate a primer extension reaction off its 3' terminus. One or both of the primers can additionally contain a 5'-terminal (5'-most) non-priming portion, i.e., a region that does not participate in hybridization to the preferred template.

5. Nucleic Acid Sequence Analysis

Nucleic acid sequence analysis is a well known procedure for determining the sequence of nucleotides and is applied to the present methods to determine the nucleotide sequence in an identifier oligonucleotide or PCR amplification product of this invention. Nucleic acid sequence analysis is approached by a combination of (a) physiochemical techniques, based on the hybridization or denaturation of a probe strand plus its complementary target, and (b) enzymatic reactions with endonucleases, ligases, and polymerases.

In assays using nucleic acid hybridization, detecting the presence of a DNA duplex in a process of the present invention can be accomplished by a variety of means.

In one approach for detecting the presence of a DNA duplex, an oligonucleotide that is hybridized in the DNA duplex includes a label or indicating group that will render the duplex detectable. Typically such labels include radioactive atoms, chemically modified nucleotide bases, and the like.

The oligonucleotide can be labeled, i.e., operatively linked to an indicating means or group, and used to detect the presence of a specific nucleotide sequence in a target template.

Radioactive elements operatively linked to or present as part of an oligonucleotide probe (labeled

hybridized to a DNA duplex.

Techniques for the separation of single stranded oligonucleotide, such as non-hybridized labeled oligonucleotide probe, from DNA duplex are well known, and typically involve the separation of single stranded from double stranded nucleic acids on the basis of their chemical properties. More often separation techniques involve the use of a heterogeneous hybridization format in which the non-hybridized probe is separated, typically by washing, from the DNA duplex that is bound to an insoluble matrix. Exemplary is the Southern blot technique, in which the matrix is a nitrocellulose sheet and the label is ^{32}P. Southern, J. Mol. Biol., 98:503 (1975).

The oligonucleotides can also be advantageously linked, typically at or near their 5'-terminus, to a solid matrix, i.e., aqueous insoluble solid support as previously described.

It is also possible to add "linking" nucleotides to the 5' or 3' end of the member oligonucleotide, and use the linking oligonucleotide to operatively link the member to the solid support.

In nucleotide hybridizing assays, the hybridization reaction mixture is maintained in the contemplated method under hybridizing conditions for a time period sufficient for the oligonucleotides having complementarity to the predetermined sequence on the template to hybridize to complementary nucleic acid sequences present in the template to form a hybridization product, i.e., a complex containing oligonucleotide and target nucleic acid.

The phrase "hybridizing conditions" and its grammatical equivalents, when used with a maintenance time period, indicates subjecting the hybridization

Where the nucleic acid containing a target sequence is in a double stranded (ds) form, it is preferred to first denature the dsDNA, as by heating or alkali treatment, prior to conducting the hybridization reaction. The denaturation of the dsDNA can be carried out prior to admixture with a oligonucleotide to be hybridized, or can be carried out after the admixture of the dsDNA with the oligonucleotide.

Effective amounts of the oligonucleotide present in the hybridization reaction admixture are generally well known and are typically expressed in terms of molar ratios between the oligonucleotide to be hybridized and the template. Preferred ratios are hybridization reaction mixtures containing equimolar amounts of the target sequence and the oligonucleotide. As is well known, deviations from equal molarity will produce hybridization reaction products, although at lower efficiency. Thus, although ratios where one component can be in as much as 100 fold molar excess relative to the other component, excesses of less than 50 fold, preferably less than 10 fold, and more preferably less than two fold are desirable in practicing the invention.

Examples

The following examples are intended to illustrate, but not limit, the present invention.

1. Preparation of Protected Amino Acids

The synthesis of a bifunctional molecule requires protected amino acids. The amino-terminus of the amino acid is protected with fluorenylmethoxycarbonyl (FMOC) and the carboxy-terminus is protected with a pentafluorophenyl ester (Opfp). The amino acids

atmosphere to form FMOC-Tyr(OTBS)-Opfp.

FMOC-Ser(OTBS)-Opfp is similarly prepared using FMOC-Ser(tert-butyl)-Opfp (Bachem) in the reaction. FMOC-Thr(OTBS)-Opfp is also prepared in this manner using FMOC-Thr(tert-butyl)-Opfp (Bachem).

FMOC-Asp(TMSE)-Opfp having a trimethylsilyl ethyl ester (TMSE) on the side chain carboxyl group of aspartic acid is prepared by first reacting one equivalent of FMOC-Asp-O-tertbutyl (Bachem) with 1.5 equivalents of 2-trimethylsilylethanol and 1.5 equivalents if dicyclocarbodiimide (DCC) in ethyl acetate for 12 hours at room temperature under inert atmosphere to form FMOC-Asp(OTMSE)-O-tertbutyl. There after the TMSE ester is reacted with an excess of formic acid at room temperature for 14 hours to hydrolyze the tertbutyl moiety and form a free carboxyl terminus in the form of FMOC-Asp(OTMSE)-COOH. The formic acid is evaporated, and the 1 equivalent of the remaining amino acid is admixed with 1.1 equivalent of pentafluorophenol (pfp; Bachem) and 1.1 equivalent of DCC for 12 hours at room temperature under inert atmosphere to form the product FMOC-Asp(TMSE)-Opfp. The product is isolated from unreacted pfp, DCC and precursor amino acid by silica gel chromatography using 10% (v/v) ethyl acetate in hexane.

FMOC-Glu(TMSE)-Opfp having a TMSE ester on the side chain carboxyl group of glutamic acid is prepared as described above to prepare FMOC-,TMSE- and pfp protected aspartic acid, except that FMOC-Glu-O-tert-butyl (Bachem) is used in place of the aspartic acid precursor, to form FMOC-Glu(TMSE)-Opfp.

retained and admixed with 2 equivalents of Iodine in tetrahydrofuran/water, 9:1, (available from Glen Research) under inert atmosphere and maintained at room temperature for 10 minutes to form oxidized solid-support coupled linker.

One equivalent of oxidized solid support-coupled linker was then admixed with 20 equivalents of acetonitrile/acetic anhydride, 88:12, (capping reagent; Glen Research) for 10 minutes at room temperature to cap any unreacted free hydroxyls present on the solid support and form capped teflon-solid support-coupled linker.

b. Controlled Pore Glass (CPG) Support

A controlled pore glass (CPG) support and linker is particularly preferred that can also be used in the present invention. The complete structure of an exemplary CPG support is shown in Figure 3, and includes (1) a sarcosine linker moiety connecting the linker to the CPG support, (2) a succinyl-aminohexanol linker that is cleavable by aqueous ammonia to release the polypeptide-linker-oligonucleotide conjugate (A-B-C) from the solid support, and (3) a serine branch-monomer having a photolabile cleavage site for releasing the oligonucleotide from the solid support. The synthesis is carried out stepwise as described below.

N-Fmoc-amidohexan-1-ol (1) was first prepared to form the aminohexanol linker moiety. To that end, 6-Amino-1-hexanol (0.75 g, 6.4 mmol) was dissolved in sat. aq. Na_2CO_3 (10 ml) and cooled on ice. 9-Fluorenylmethyl chloroformate (Fmoc-Cl, 1.83 g, 7.1 mmol) in THF (25 ml, freshly distilled) was added slowly under vigorous stirring. The solution is acidified with 10% (w/v) citric acid, extracted with

added with stirring at room temperature under an inert atmosphere (Ar). The reaction mixture was stirred over night, evaporated to dryness, redissolved in chloroform (4 x 25 ml) and extracted with sat. aq. $NaHCO_3$ (50 ml). The aq. phase was back-extracted once with chloroform (25 ml) and the combined organic phases were dried from Na_2SO_4, filtered, and evaporated to yield the crude product as a brownish oil. The oil was dissolved in chloroform/ethylacetate and triturated once with hexane and then purified by silica gel column chromatography (150 ml silica) using $CHCl_3$/MeOH/DIPEA, 94/5/1, as the mobile phase. The fractions were analyzed by tlc and the appropriate fractions were pooled and evaporated to give a white foam designated compound 3. The yield of compound 3 was 4.29 g, and the
MS (EI, *m/e*) was 630 (calcd for $C_{39}H_{35}NO_7$ + H 630).

Controlled pore glass (CPG) was activated for coupling to the above prepared linkers. To that end, CPG (Sigma G5019, aminopropyl-CPG, 4.97 g) was suspended in trichloroacetic acid in dichloromethane (3%, 20 ml) and agitated in a shaker for 4 hours. The CPG was isolated by filtration, washed three times with dichloromethane (20 ml), three times with DIPEA in chloroform (10%, 20 ml) and three times with diethyl ether to form activated CPG. The activated CPG was dried *in vacuo*.

Bi-functional CPG [bf-CPG, CPG-Sar-suc-aho-Ser(*O*-DMT)(*N*-Fmoc)] was then prepared comprising all of the element shown in Figure 3.

i. Activated CPG (2.0 g, loading 83 μmole/g) was placed in a filter funnel (25 ml). Fmoc-Sar (313.5 mg, mmol) was dissolved in DMF (2.8 ml), mixed with pyBOP (525.9 mg, mmol) in DMF (2.8 ml) and the combined, activated amino acid mixture was

to light of about 350 nm and which upon cleavage results in the release of soluble polypeptide. Cleavage can be regulated by the time and intensity of irradiation using, for example, a Rayonet RPR Reactor as described by Barany et al., supra.

4. Coupling of Nucleotide to Solid Support

a. Deprotect DMT

The capped teflon solid support-coupled linker prepared in Example 3 was admixed with 3% DCA in DCM for 10 minutes at room temperature under inert atmospheres to remove the DMT protecting group from the linker and form a free hydroxyl group. The deprotected linker/support was then washed 3 times with DCM. The deprotected linker/support is ready for addition of nucleotide.

b. Addition of Nucleotide to Teflon Support

One equivalent of deprotected linker/support was admixed with about 20 equivalents of a desired blocked nucleotide phosphoramidite and 20 equivalents of tetrazole (0.45M in acetonitrile) to form a coupled nucleotide/linker/support (coupled nucleotide complex). The coupled nucleotide complex was then washed with an excess of acetonitrile to remove unreacted reagents. All blocked nucleotide phosphoramidites were obtained from Glen Research and contain a DMT protected blocked 5'hydroxyl, a cyanoethyl ester (CNE) and a diisopropylamine group at the 3'-phosphoramidite. In addition, the adenine and cytosine derivatives contained a benzoyl group on the base's free nitrogen and the guanosine derivative contains an isobutyl group on the 2-amino group of the purine base.

One equivalent of coupled nucleotide complex

was then admixed with 2 equivalents of Iodine in tetrahydrofuran/water, 9:1, for oxidation as before in Example 3 to oxidize the coupled nucleotide complex.

Thereafter, one equivalent of oxidized nucleotide complex was admixed with 20 equivalents of capping reagent as before in Example 3 to cap any unreacted free hydroxyls and to form capped solid phase-coupled nucleotide complex.

c. Coupling of Nucleotide to Bifunctional CPG

As a demonstration of coupling chemistry for addition of a nucleotide residue to the bifunctional CPG support, we have synthesized and coupled an oligodeoxyribonucleotide onto bf-CPG. To that end, oligonucleotides were synthesized on a ABI 394 DNA synthesizer using the standard 1 μmole scale synthesis cycle and commercially available reagents and phosphoramidites. The solid support used was the bf-CPG (20 mg, 1 μmol) placed in a commercially available empty synthesis column for the ABI synthesizers. Repetitive yield were calculated to 98.9% as judged from collecting the detritylation, deluting in *p*-toluene sulfonic acid monohydrate in acetonitril (0.1 M) and measuring the absorptions at 498 nm.

The oligonucleotide was released from the support by conc, aq. ammonia (shaking for more than 24 hours). The oligonucleotide was isolated using commercially available OPC-cartridges and analyzed by HPLC and PAGE (radiolabeled with ^{32}P-g-ATP and T4-kinase).

5. Coupling of Amino Acid to Solid Support

a. Deprotect Nucleotide Complex

One equivalent of capped teflon solid phase-coupled nucleotide complex was admixed with 1

_____ device. Reagents used were the commercially available Fmoc amino acid derivatives either as O-pfp esters (for Val and His[Fmoc]) or the free acid acivated with pyBOP (for Phe, Gln and Pro). Coupling was conducted in the automated peptide synthesizer _____ using conventional _____ chemistry. The success of the coupling reaction was monitored by a positive and a negative Kaiser test after the Fmoc deprotection and the coupling steps, respectively.

The peptide was deprotected and released from an aliquot of the bf-CPG by treatment with *i*. piperidine in DMF (2/8, 2 x 5 minutes) *ii*. conc, aq. ammonia (shaking for 24 hours). The peptide was analyzed by reversed phase HPLC as one major peak. MS (Ion Spray, *m/e*) 813 (calcd for $C_{39}H_{60}N_{10}O_9$ + H 813).

6. Elongation of the Conjugate

The conjugate can be lengthened by alternating cycles of addition of nucleotides and amino acids. The following alternating cycles are repeated until the conjugate has desired length amino acid polymer and oligonucleotide polymer.

a. Addition of Nucleotides

To couple an additional nucleotide, the 5'-OH on the terminal nucleotide is deprotected with DCA following the protocol described previously in Example 4A for the deprotection of the teflon linker/support. Thereafter, a protected nucleotide is added as described in Example 4B.

b. Addition of Amino Acids

To couple an additional amino acid, the amino-FMOC terminus of the last amino acid added to

(0.1 M) and measuring the absorptions at 498 nm.

The 45-mer oligonucleotide above has two flanking 15-mer regions that provide a site for PCR primers, and a middle 15-mer that is the coding sequence for the peptide using (arbitrarily) the commaless codons CTA for F, ATC for H, ACC for P, ACA for Q and GCG for V.

The peptide-oligonucleotide conjugates formed were realesed from the support by conc, aq. ammonia (shaking for more than 24 hours). The peptide-oligonucleotide conjugates were isolated using commercially available OPC-cartridges and analyzed by HPLC and PAGE (radiolabeled with ^{32}P-g-ATP and T4-kinase).

d. Synthesis of a Peptide-Oligonucleotide Conjugate Library on a bf-CPG Support

Using a bifunctional (bf) CPG support as described in Example 3B, a peptide-oligonucleotide conjugate library is constructed as described below in the following steps:

Step 1. The solid support, bf-CPG, is placed in a commercially available empty synthesis column for use in an ABI 394 DNA synthesizer (Applied Biosystems, Inc., Foster City, CA). The 3'- oligonucleotide primer sequence (SEQ ID NO ___) is synthesized and then chemically attached to the bf-CPG support according to standard manufacturer's procedures using the standard synthesis cycles and commercially available reagents and phosphoramidites.

Step 2. The support after Step 2 is divided into a number of aliquots representing the number of different monomer amino acids to be added at library position 1 of the polypeptides to be synthesized. Each Fmoc-amino acid residue Monomer(1) is coupled

CPG conjugate.

Step 4. The different Monomer(1)-functionalized bf-CPG's are pooled, mixed and divided into a number of aliquots representing the number of different amino acid residue monomers to be added in the library at position 2. Monomer(2) is attached to the CPG as described under Step 2 for Monomer(1).

Step 5. The oligonucleotide sequence that is the unit identifier for the appropriate Monomer(2) is attached to each of the Monomer(1)-functionalized bf-CPG's by placing each of the different Monomer(1)-functionalized bf-CPG's in separate commercially available empty synthesis columns for the DNA synthesizers.

These steps, Steps 4-5, are repeated until the appropriate number of monomers and corresponding unit identifier oligonucleotides have been incorporated.

Step 6. The combined solid supports are placed in a commercially available empty synthesis column for the DNA synthesizers. The 5'-oligonucleotide primer sequence is chemically attached to the Monomer(n)-Monomer(1)-functionalized bf-CPG as described previously using the standard synthesis cycles and commercially available reagents and phosphoramidites. The final DMT-group is left attached to the oligonucleotide-peptide conjugates to be used as an affinity tag in the purification step.

Step 7. The combinatorial peptide-oligonucleotide library is deprotected and released from the support by treatment first with TBAF for the appropriate time, and then with conc, aq. ammonia (shaking for more than 24 hours). The peptide-oligonucleotide conjugates were isolated and purified using commercially available OPC-cartridges

ii. Removal of the Bz Group

The conjugate is admixed with an excess of aqueous ammonia and maintained at 60°C overnight under inert atmosphere to remove the benzyl (Bz) group protecting the side chain amino group of lysine.

iii. Removal of the MTr, Bum and Tpm Groups

The conjugate is admixed with 20 to 50 percent TFA and maintained at room temperature for about 5 minutes under inert atmosphere to remove the MTr, Bum or Tpm groups protecting the side chains of arginine, histidine or cysteine, respectively. Thereafter, the conjugate is neutralized with triethanolamine and $CHCl_3$.

iv. Removal of the Formyl Group

The conjugate is admixed with aqueous buffer at pH 12 and maintained at room temperature for about 5 minutes under inert atmosphere to remove the formyl group protecting the reactive 2-amino group of tryptophan.

8. Cleavage of Conjugate from Solid Support

a. Cleavage of the Teflon Support

After the protecting groups are removed from the conjugate on the teflon support, the bifunctional molecule is removed from the solid support by admixing the conjugate with a cleaving solution of 100 mM sodium periodate, 100 mM sodium phosphate buffer, pH 7.2, in acetonitrile/water (1:4 v/v). The admixture is maintained with agitation at room temperature with exclusion of light. After 4 hours of agitation, the liquid phase removed and the solid support is washed with excess water and methanol. The wash solutions are then removed and 1 umole of solid support are

9. Preparation of a Library of Bifunctional Molecules

Using the synthetic procedures of Examples 1-8, the methods for producing a bifunctional molecule are detailed. To form a library of molecules, additional manipulations are required. First, the synthesis is conducted including the steps of aliquoting, adding different units to each aliquot, and pooling the aliquots to sequentially build the library. Second, if desired, the PCR primer binding sites and the unit identifier oligonucleotides can be added as presynthesized blocks rather than added nucleotide by nucleotide.

a. Synthesis of Protected Oligonucleotides

Using this procedure, PCR primer binding site oligonucleotides P1 and P2 were synthesized having the nucleotide sequences shown in Table 2, but having an DMT at the oligonucleotide's 5' terminus, and having a CNE ester and an amino diisopropyl phosphoramidate at the oligonucleotide's 3' terminus. Similarly, unit identifier oligonucleotides were synthesized for glycine (Gly) and methionine (Met) having 6 nucleotides per unit and having the blocked termini described above. The unit identifier oligonucleotide sequences are shown in Table 2.

TABLE 2

Designation	Oligonucleotide Sequence
P1	5'-GGGCCCTATTCTTAG-3'
P2	5'-AGCTACTTCCCAAGG-3'
Z^{gly}	5'-CTCATG-3'
Z^{met}	5'-ACGGTA-3'

ACGGTACTCATG-**P1**-LINK-gly.met , and
ACGGTAACGGTA-**P1**-LINK-met.met .

The two aliquots are then pooled to form a mixture of the four bifunctional molecules.

In Step 3, the pool from Step 2 is divided into two aliquots. The first aliquot is subjected to a sequential coupling as before, adding glycine and the oligonucleotide Z^{gly} . Thereafter, protected oligonucleotide P2 is coupled to the growing bifunctional molecules in the pool as described for a single protected nucleotide in Example 4 to form the structures:

P2CTCATGCTCATGCTCATG**P1**-LINK-gly.gly.gly ,
P2CTCATGCTCATGACGGTA**P1**-LINK-met.gly.gly ,
P2CTCATGACGGTACTCATG**P1**-LINK-gly.met.gly ,and
P2CTCATGACGGTAACGGTA**P1**-LINK-met.met.gly .

The second aliquot is subjected to a sequential coupling as before, adding methionine and the oligonucleotide Z^{met}. Thereafter, protected oligonucleotide P2 is coupled to the growing bifunctional molecules in the pool as described for a single protected nucleotide in Example 4 to form the structures:

P2ACGGTACTCATGCTCATG**P1**-LINK-gly.gly.met ,
P2ACGGTACTCATGACGGTA**P1**-LINK-met.gly.met ,
P2ACGGTAACGGTACTCATG**P1**-LINK-gly.met.met ,and
P2ACGGTAACGGTAACGGTA**P1**-LINK-met.met.met .

The two aliquots are then pooled to form a mixture of the eight bifunctional molecules.

The resulting pool of eight different bifunctional molecules represents a small library produced according to the methods of this invention. By increasing the alphabet size one increases the number of aliquots per step.

What Is Claimed Is:

1. A bifunctional molecule according to the formula A-B-C, wherein A is a chemical moiety, B is a linker molecule operatively linked to A and C, and C is an identifier oligonucleotide comprising a sequence of nucleotides which sequence identifies the structure of chemical moiety A.

2. The bifunctional molecule of claim 1 wherein A is a polymer comprising a linear series of chemical units represented by the formula $(X_n)_a$, wherein X is a single chemical unit in polymer A; and identifier oligonucleotide C is represented by the formula $(Z_n)_a$, wherein Z is a unit identifier nucleotide sequence within oligonucleotide C that identifies the chemical unit X at position n; and wherein

n is a position identifier for both X in polymer A and Z in oligonucleotide C having the value of 1+i. where i is an integer from 0 to 10, such that when n is 1, X or Z is located most proximal to the linker, and a is an integer from 4 to 50.

3. The bifunctional molecule of claim 2 wherein said unit identifier nucleotide sequence Z has a length of from 2 to 8 nucleotides.

4. The bifunctional molecule of claim 1 wherein said chemical moiety A is an oligosaccharide, polypeptide, glycolipid, lipid, proteoglycan, glycopeptide or oligonucleotide.

5. The bifunctional molecule of claim 2 wherein said polymer A is a polypeptide, X is an amino acid residue in said polypeptide, and unit identifier nucleotide sequence Z is a hexanucleotide sequence that identifies the amino acid residue at position n in polypeptide A.

6. The bifunctional molecule of claim 5 wherein

16. The library of claim 12 wherein X is an amino acid and said unit identifier nucleotide sequence Z has a length from 3 to 6 nucleotides.

17. The library of claim 11 wherein each of said species of bifunctional molecules in said plurality is present in molar equivalents of from 0.2 to 10.0.

18. The library of claim 12 wherein said identifier oligonucleotide C in each of said species of bifunctional molecules has a nucleotide sequence according to the formula $P1-(Z_n)_s-P2$, where P1 and P2 are nucleotide sequences that provide PCR primer binding sites adapted to amplify the identifier oligonucleotide, and where the nucleotide sequences of P1 and P2 are shared by all bifunctional molecule species in the library.

19. A method for identifying a chemical structure that participates in a preselected binding interaction with a biologically active molecule, said chemical structure being present in a library of bifunctional molecules according to claim 11, comprising the steps of:

a) admixing in solution said library of bifunctional molecules with the biologically active molecule under binding conditions for a time period sufficient to form a binding reaction complex;

b) isolating the complex formed in step (a); and

c) determining the nucleotide sequence of the identifier oligonucleotide in the isolated complex and thereby identifying the chemical structure that participated in the preselected binding interaction.

20. The method of claim 19 wherein said biologically active molecule is affixed to a solid support.

21. The method of claim 19 wherein said

to add an additional chemical unit X and corresponding identifier oligonucleotide Z to the bifunctional molecules in the admixture.

26. The method of claim 25 wherein said repetition of steps (c) and (d) are repeated on the admixture from 1 to 6 times, thereby forming a polymer A on said bifunctional molecules such that a is from 3 to 10.

27. The method of claim 19 wherein said linker molecule is a bifunctional solid support selected from the group consisting of bf-CPG or *o*-NB-bf-CPG.

28. The method of claim 24 wherein said linker molecule is a bifunctional solid support selected from the group consisting of bf-CPG or *o*-NB-bf-CPG.

29. A bifunctional solid support for synthesizing an oligopeptide/ oligonucleotide conjugate comprising:

a solid support, said solid support being of a type which is dispersible in aqueous solution,

a first linkage unit coupled to said solid support,

a second linkage unit coupled to said first linkage unit, and

a bifunctional unit coupled to said second linkage unit,

said bifunctional unit having a first leaving group employable for oligopeptide synthesis and a second leaving group employable for oligonucleotide synthesis,

said first leaving group being N-FMOC or its functional equivalent,

said second leaving group being O-DMT or its functional equivalent,

said second linkage unit being coupled to said first linkage unit by means of a bond cleavable

34. A bifunctional solid support as described in claim 33 wherein:

said solid support is aminopropyl-CPG,

said first linkage unit includes a sarcosine linker coupled to the aminopropyl-CPG and a succinyl linker coupled to the sarcosine linker by means of an amide bond,

said second linkage unit includes an aminohexanol group coupled to said succinyl linker by means of an alkyl ester,

said bifunctional unit includes an L-serine residue, the amino end of the serine being coupled by means of an amide bond to said aminohexanol linker, the carboxyl end of said serine being coupled to an FMOC leaving group, the hydroxyl end of said serine being coupled to an O-DMT leaving group.

35. A bifunctional solid support as described in claim 29 further comprising:

a third linkage unit interposed between and coupled to said bifunctional unit and said first leaving group,

said third linkage unit being cleavable by exposure to ultra- violet light.

36. A bifunctional solid support as described in claim 35 wherein said third linkage unit includes a 3-nitro-4-O-ethyl benzoate group coupled by means of an amide bond to the amino end of said serine and coupled by means of an ester bond to an FMOC blocked amino acid.

37. A bifunctional solid support as described in claim 36 wherein:

said solid support is aminopropyl-CPG,

said first linkage unit includes a sarcosine linker coupled to the aminopropyl-CPG and a succinyl linker coupled to the sarcosine linker by means of an

ultraviolet light.

41. An element of a library of oligopeptide/oligonucleotide conjugates comprising:

a bifunctional unit,

an oligopeptide attached to said bifunctional unit, and

an oligonucleotide attached to said bifunctional unit.

42. An element of a library of oligopeptide/oligonucleotide conjugates as described in claim 41 further comprising a cleavable bond for coupling said bifunctional unit to said oligopeptide, said cleavable bond being cleavable by exposure to ultraviolet light.

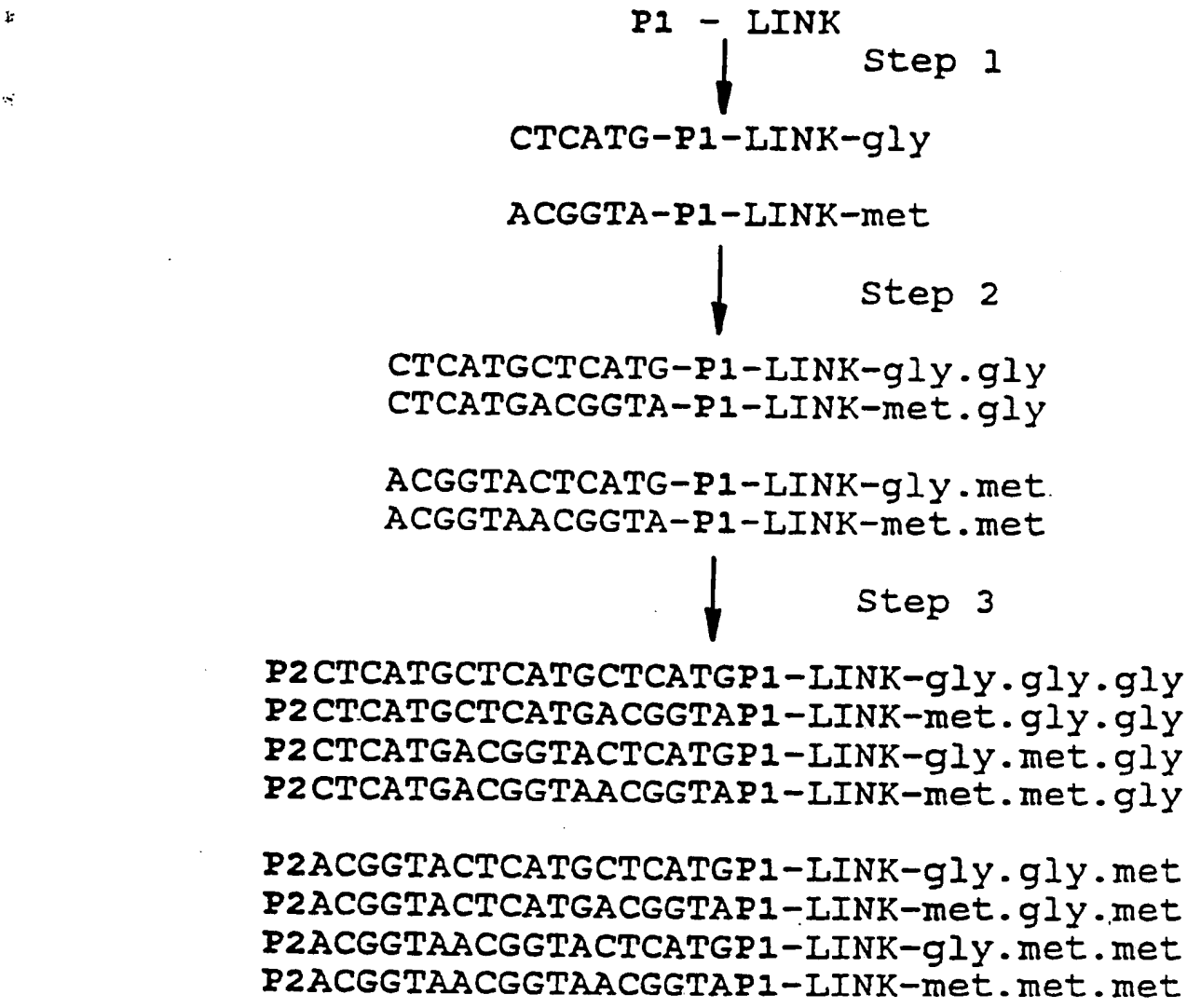

FIG. 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/03127

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C12Q 1/70; CO7K 5/00, 13/00; G01N 33/53
US CL :435/6; 530/300, 350; 435/7.1, 500
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/91; 536/24.2, 24.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	NATURE, VOLUME 354, ISSUED 07 NOVEMBER 1991, LAM ET AL., "A NEW TYPE OF SYNTHETIC PEPTIDE LIBRARY FOR IDENTIFYING LIGAND-BINDING", PAGES 82-84, SEE ENTIRE DOCUMENT.	1-6, 11-22

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 17 MAY 1993	Date of mailing of the international search report 16 JUN 1993
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C: 20231 Facsimile No. NOT APPLICABLE	Authorized officer EGGERTON CAMPBELL Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

International application No.
PCT/US93/03127

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

USPTO, APS (OLIGOPEPTID?, OLIGONUCLEOTID?, LIBRAR?, COMBINATOR?, SOLID (W) SUPPORT)
DIALOG (OLIGONUCLEOTID?, OLIGOPEPTID?, SYNTHES?, CLEAV?, SOLID (W) SUPPORT)